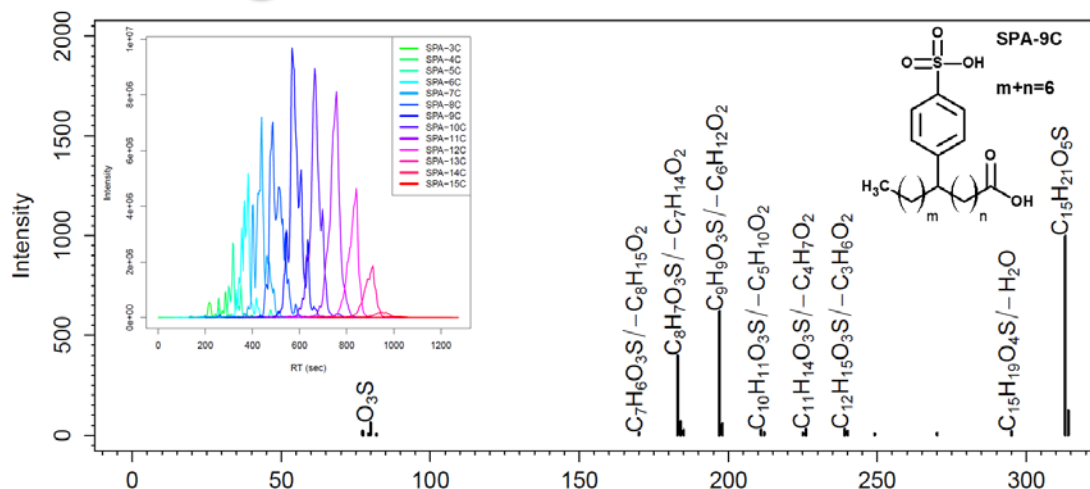


Adding Complex Expert Knowledge into Chemical Databases: Transforming Surfactants in Wastewater



Emma Schymanski

Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg.

Email: emma.schymanski@uni.lu

Juliane Hollender (Eawag, Dübendorf, Switzerland)

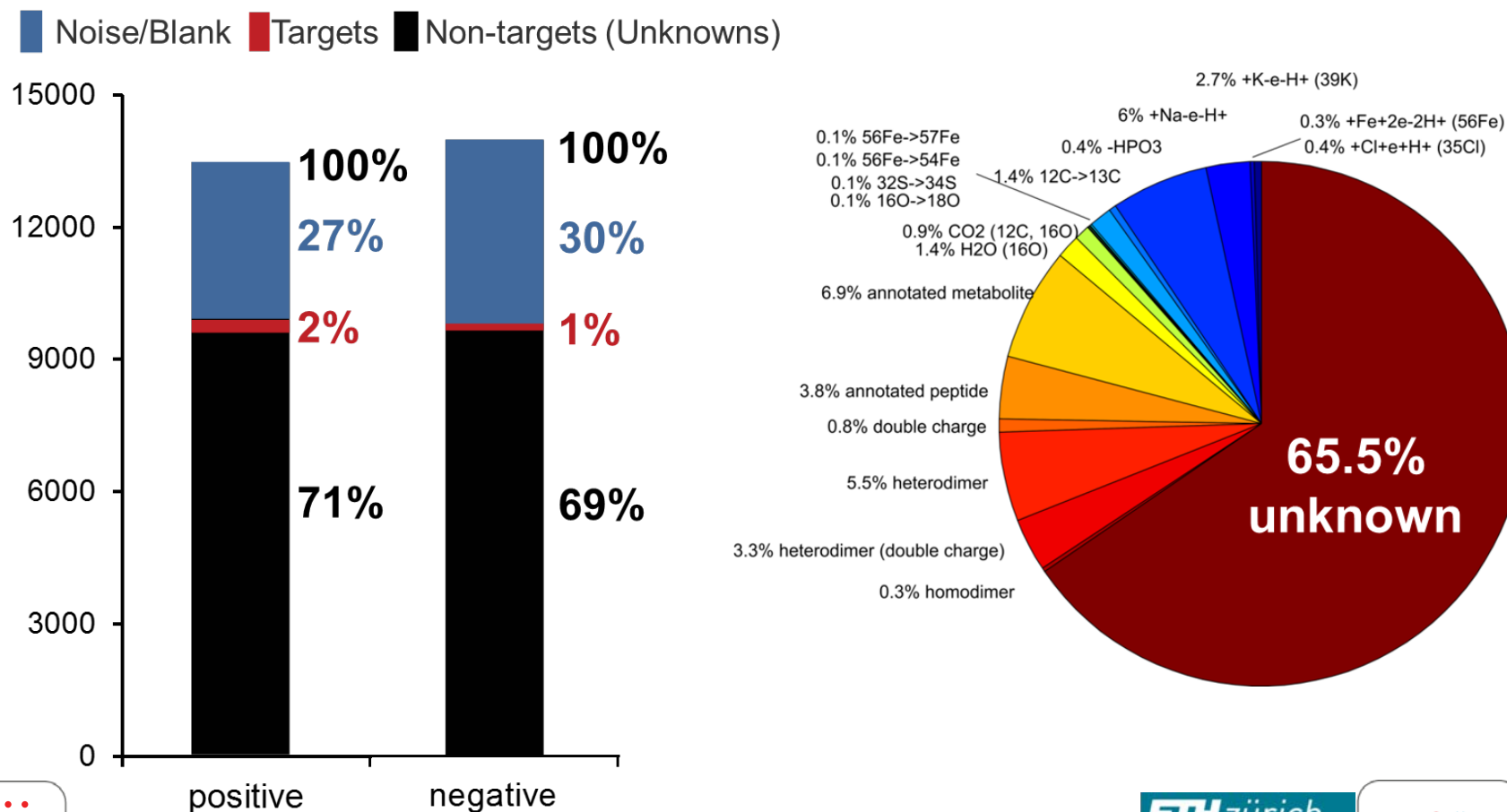
Chris Grulke (NCCT, US EPA, Research Triangle Park, NC, USA)

Antony J. Williams (NCCT, US EPA, Research Triangle Park, NC, USA)

Why Add Complex Expert Knowledge to Databases?

Unknowns and High Resolution Mass Spectrometry

Over 60 % of HR-MS peaks are potentially *relevant* but *unknown*

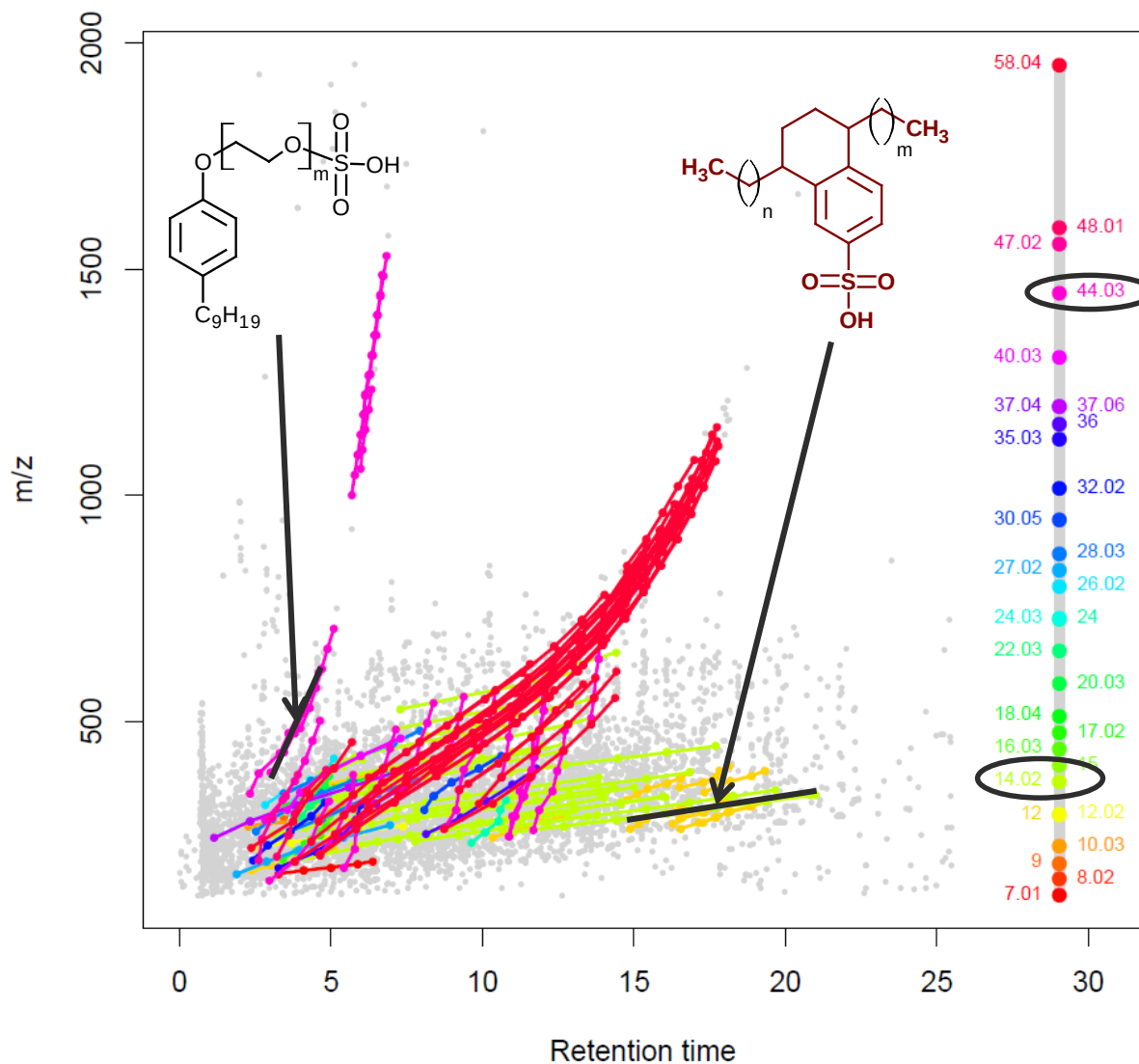


Why Add Complex Expert Knowledge to Databases?

Homologous Series and UVCBs in our samples

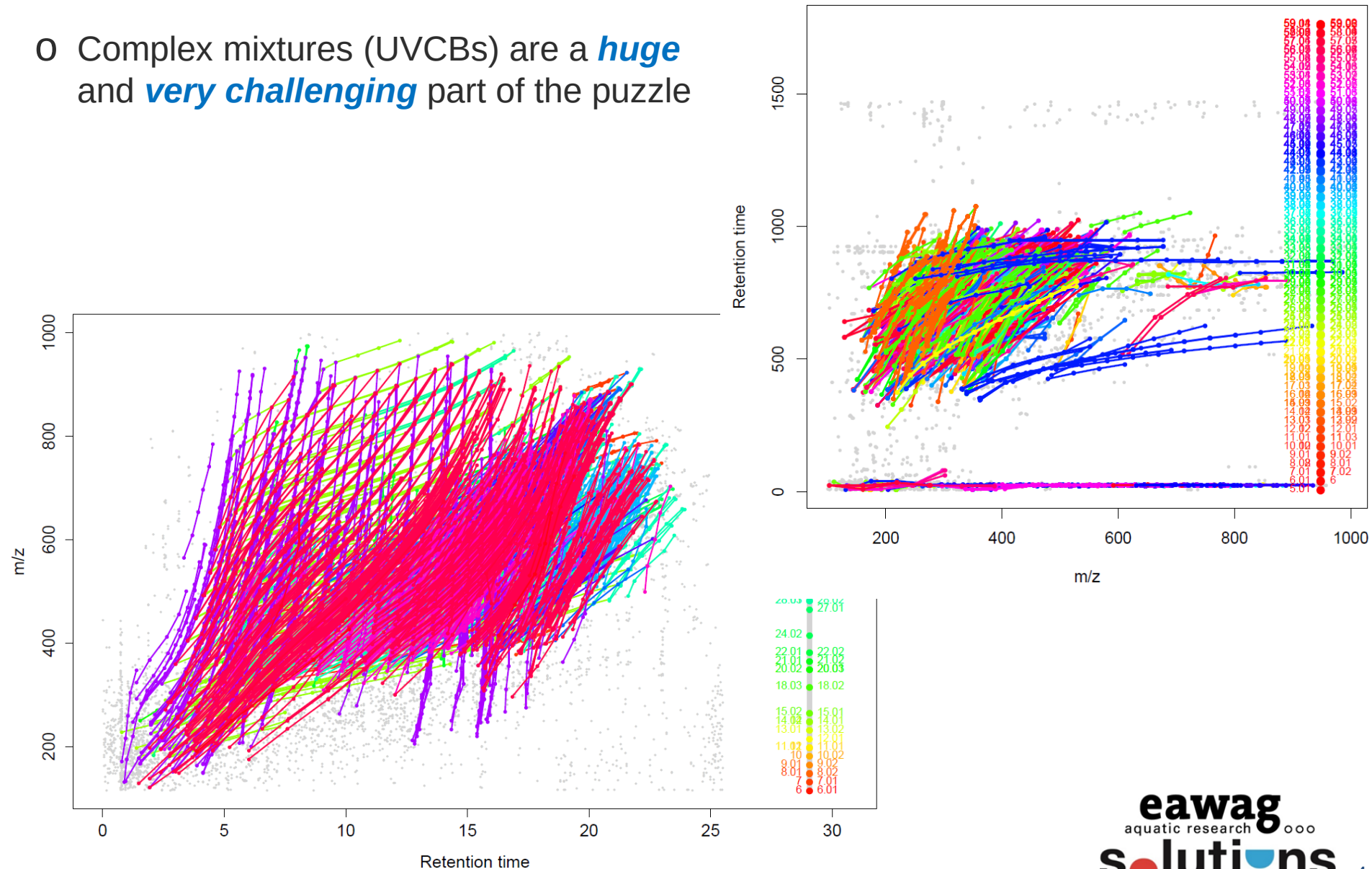
eawag
aquatic research

nontarget

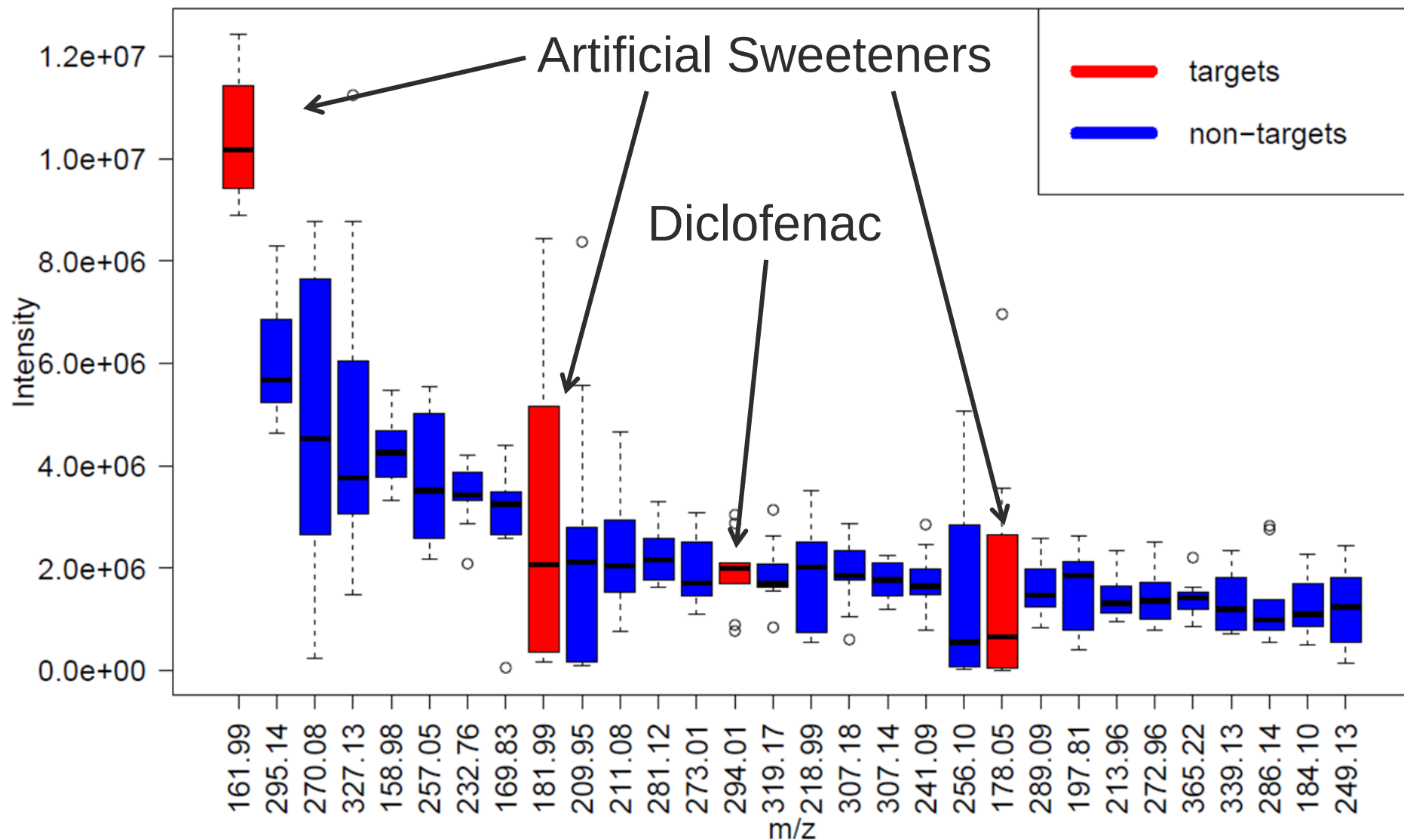


Why Add Complex Expert Knowledge to Databases?

- Complex mixtures (UVCBs) are a **huge** and **very challenging** part of the puzzle



Swiss Wastewater: Top 30 Peaks (ESI-)



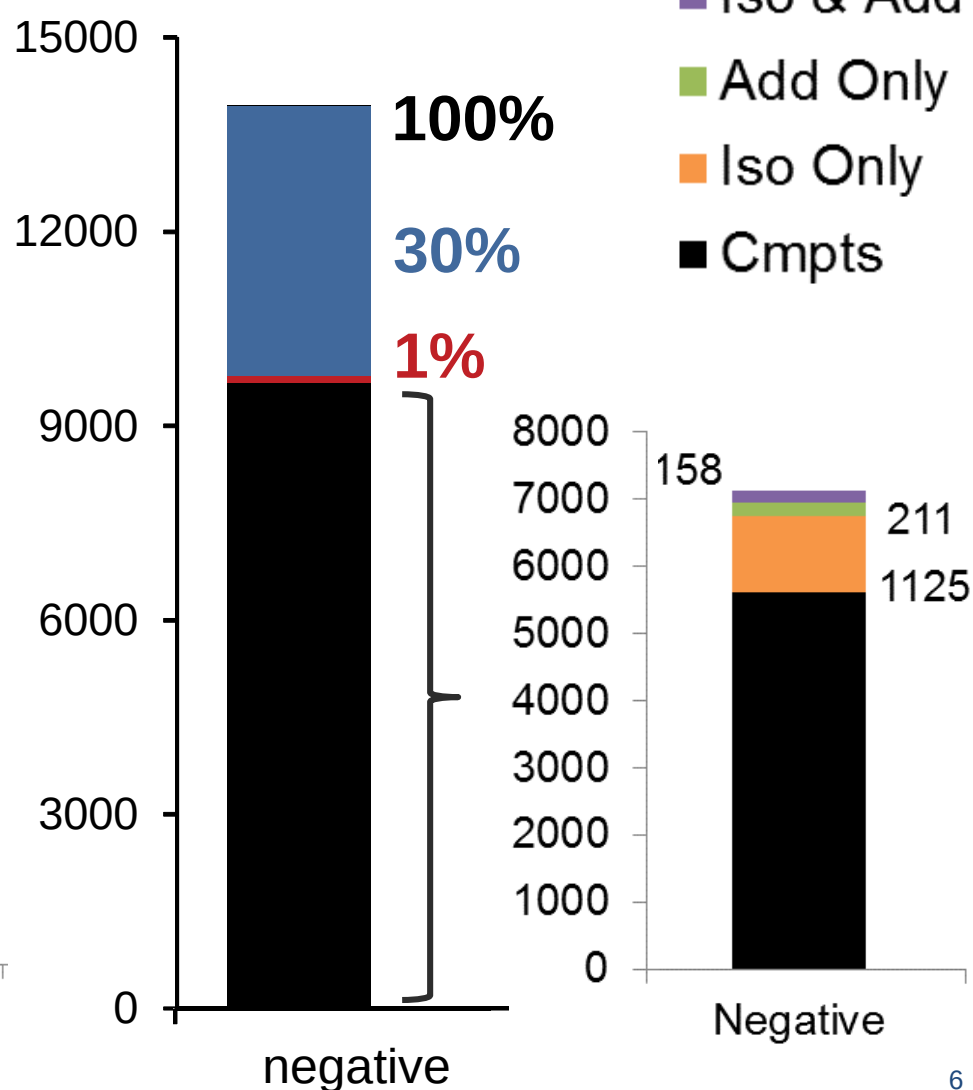
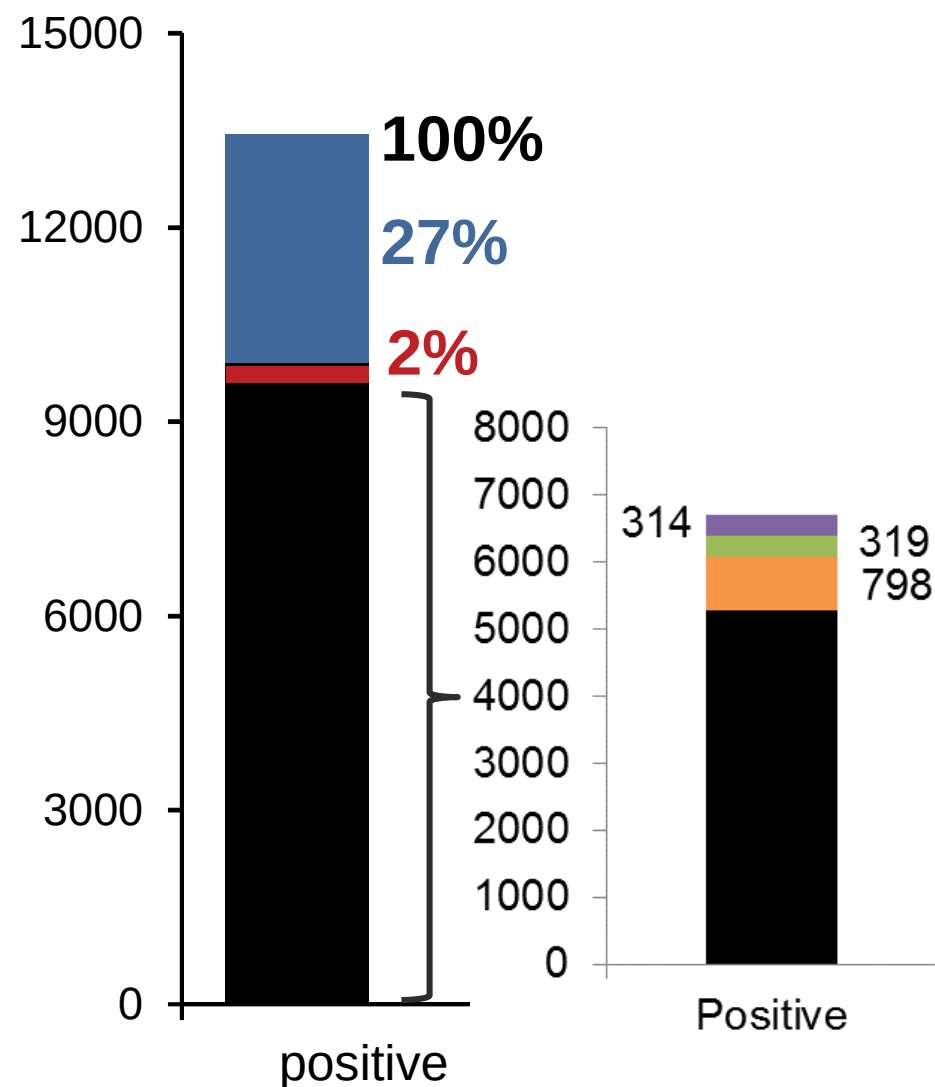
Grouping Isotopes and Adducts



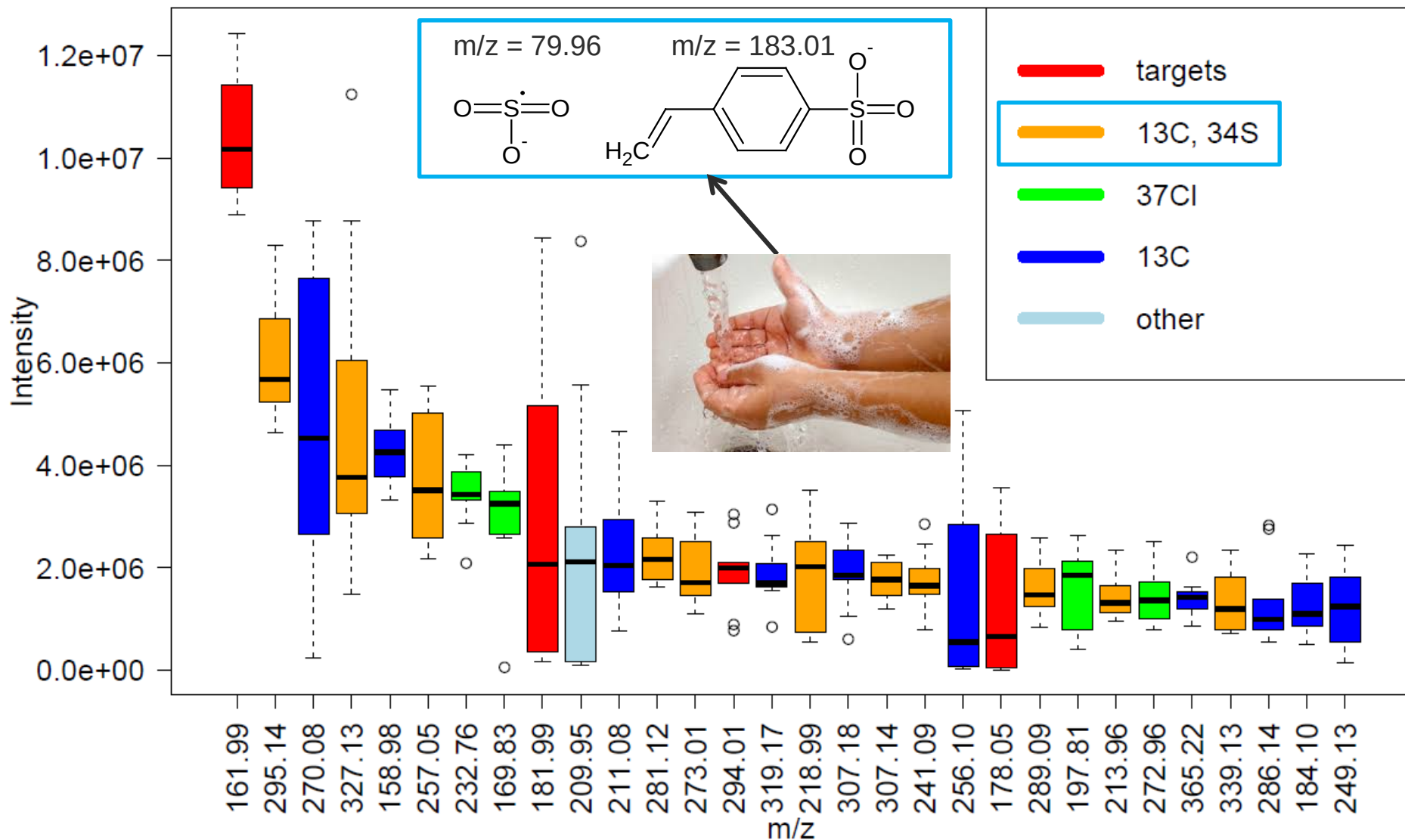
enviPat Web **nontarget**

Noise/Blank Targets Non-targets

Iso & Add
Add Only
Iso Only
Cmpts

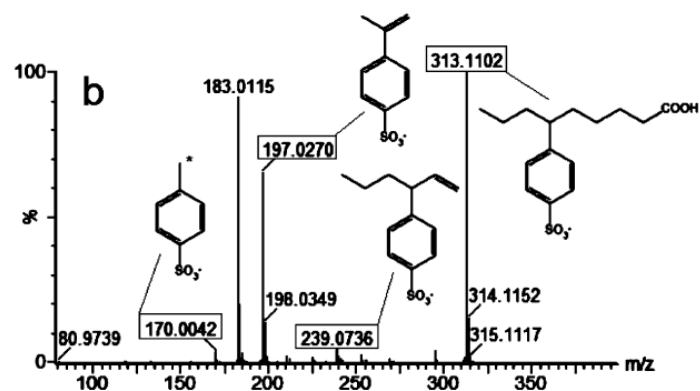
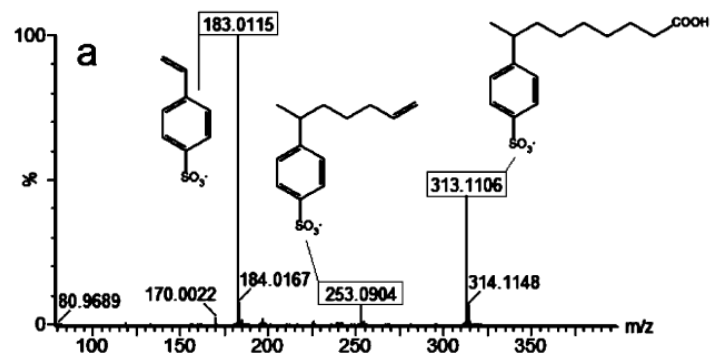


Swiss Wastewater: Top 30 Peaks (ESI-)

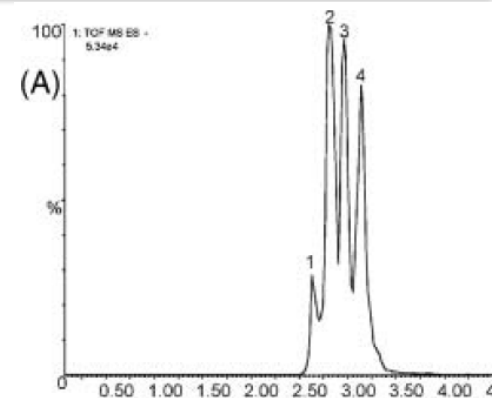


Literature sources

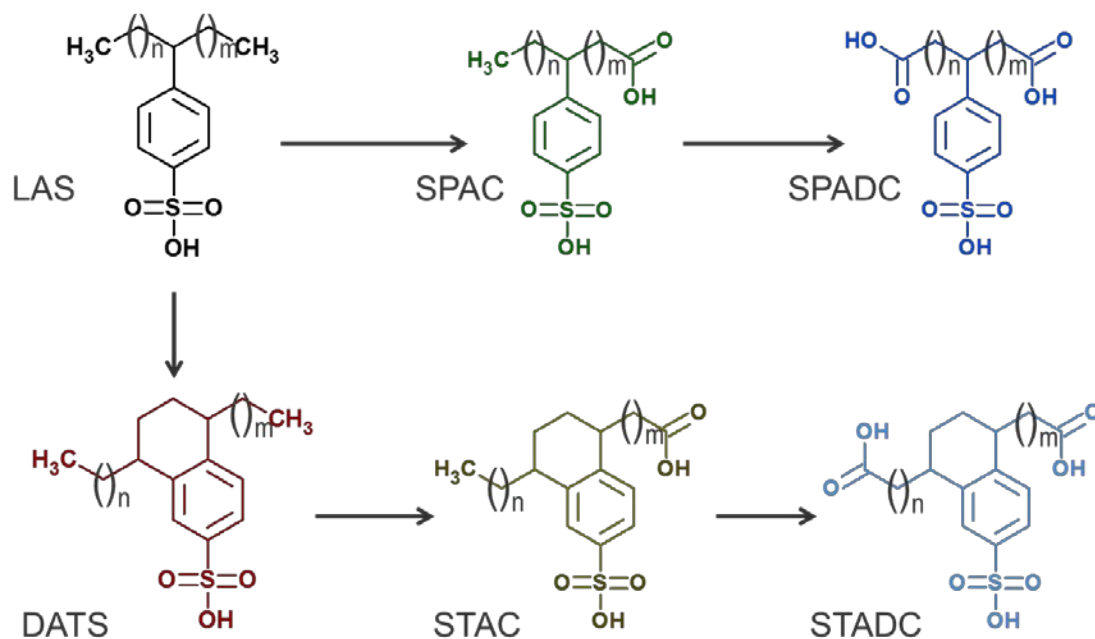
- Formulas, masses (ions), retention times and intensities
- Spectra of selected compounds (different instruments)



Lara-Martin et al. EST. 2010, 44: 1670-1676



Gonzalez et al. Rapid Comm. Mass Spec. 2008, 22: 1445-54

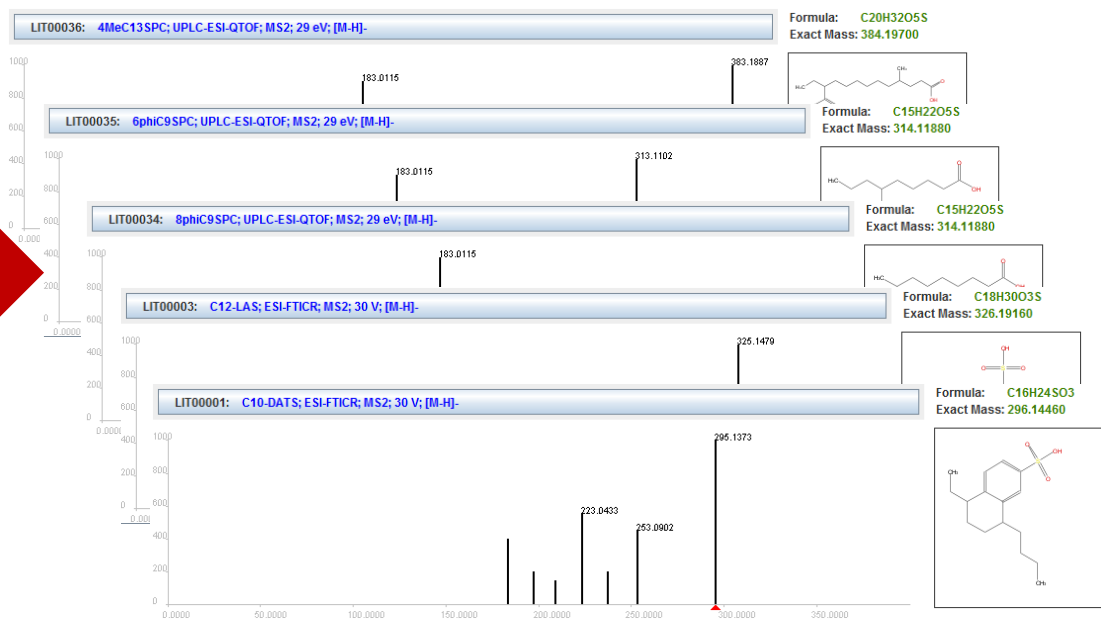
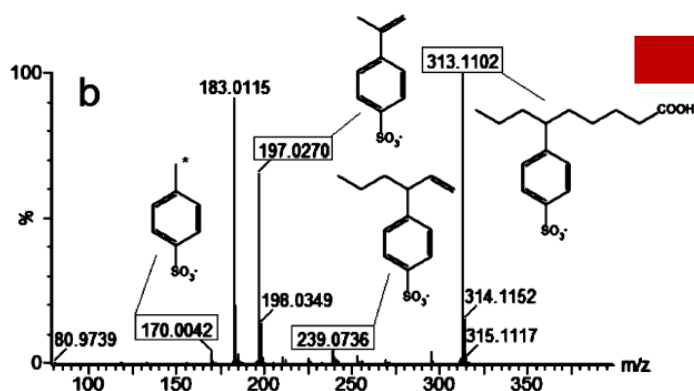
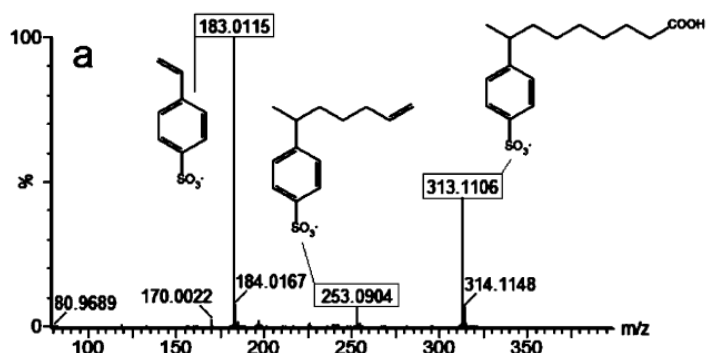


Enable access and comparison

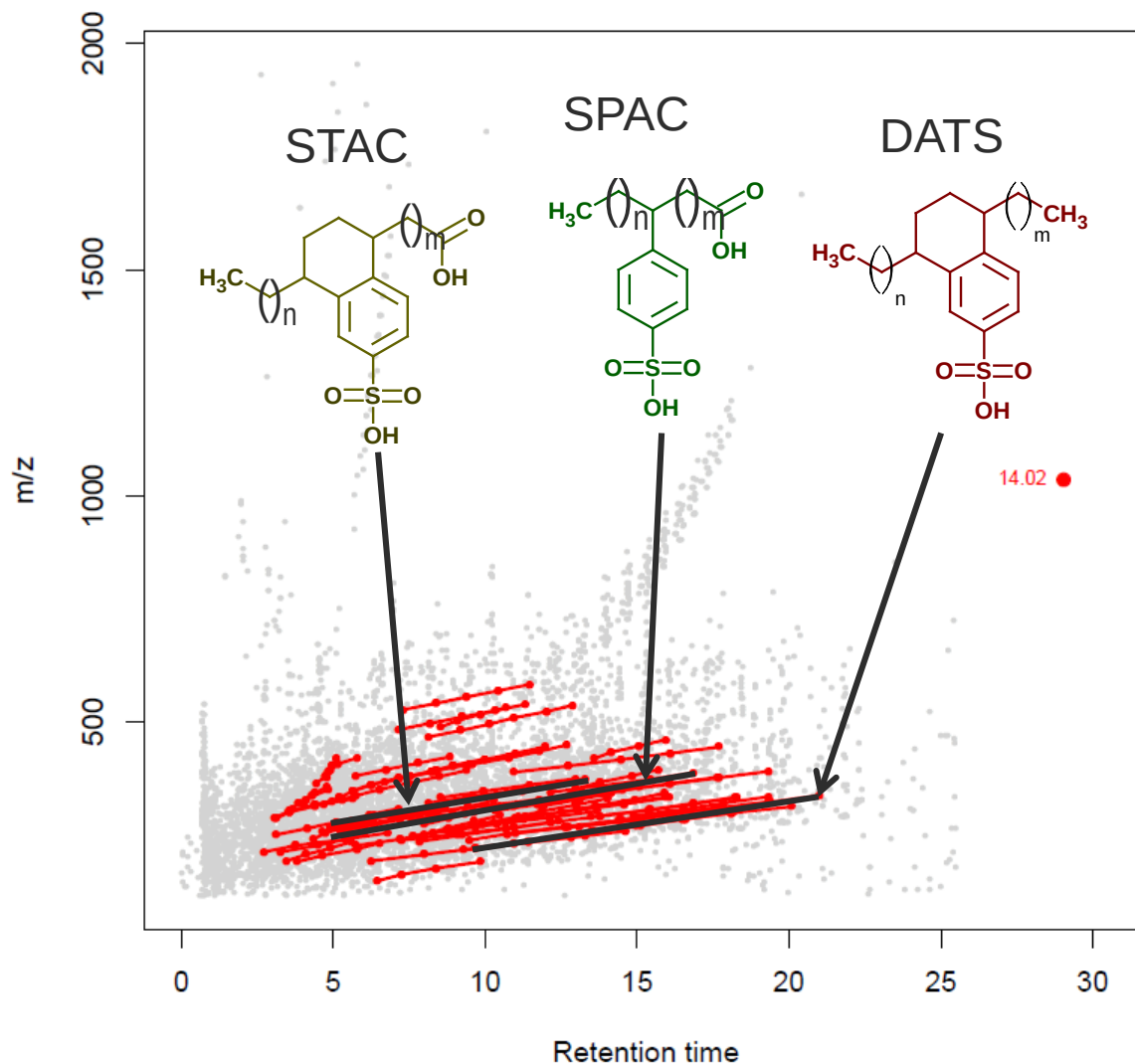
Adding literature spectra to
www.massbank.eu/MassBank/



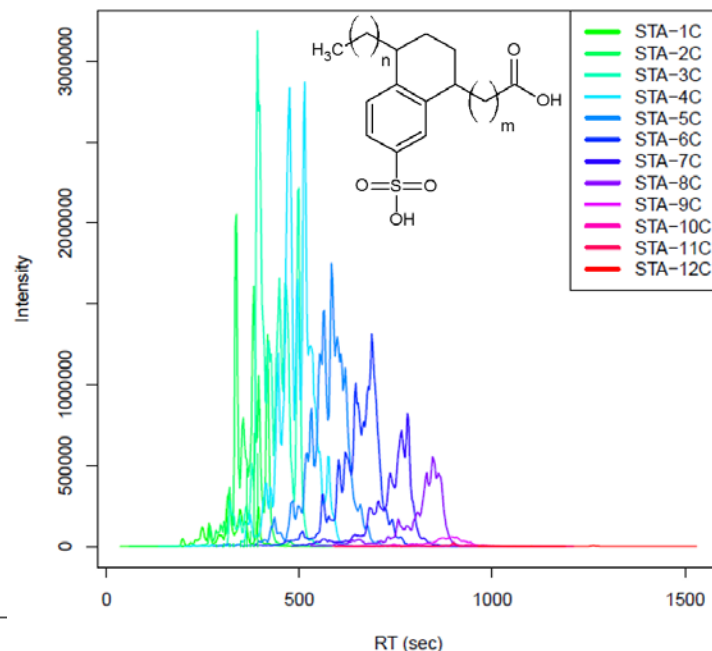
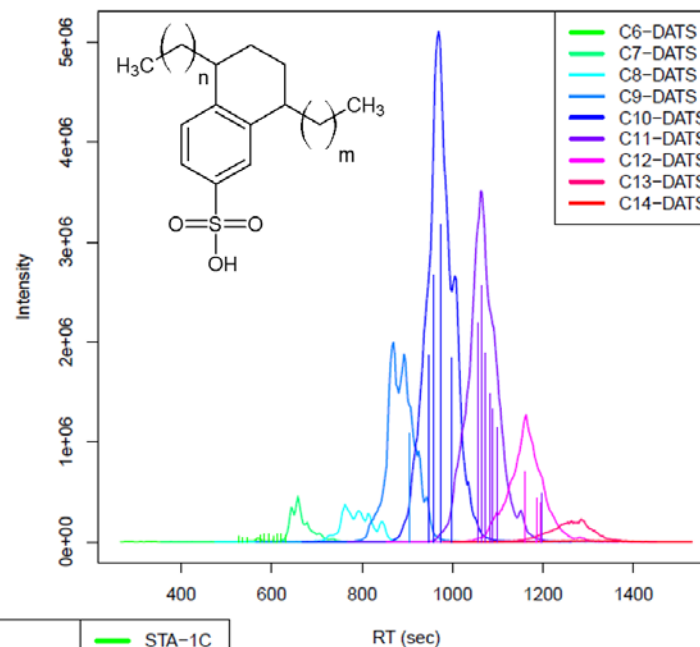
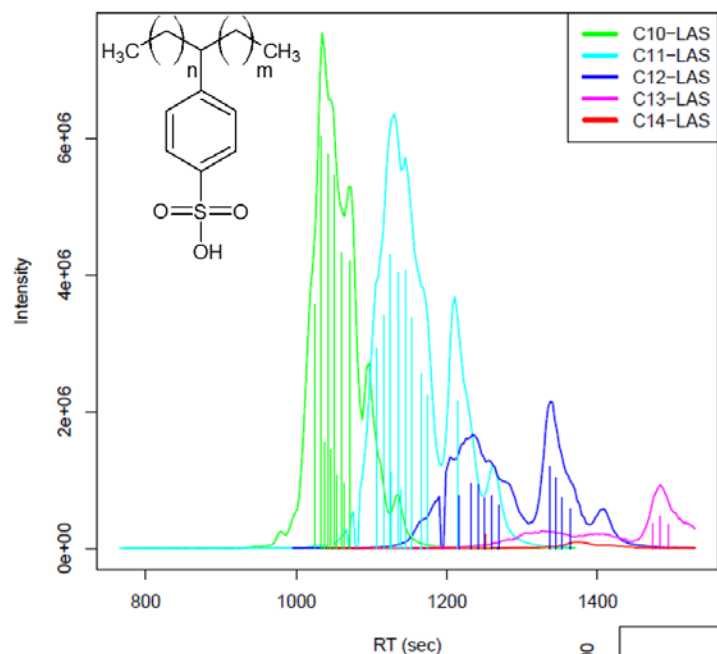
39 literature spectra (so far)



Lara-Martin et al. EST. 2010, 44: 1670-1676



Supporting Evidence for Homologues

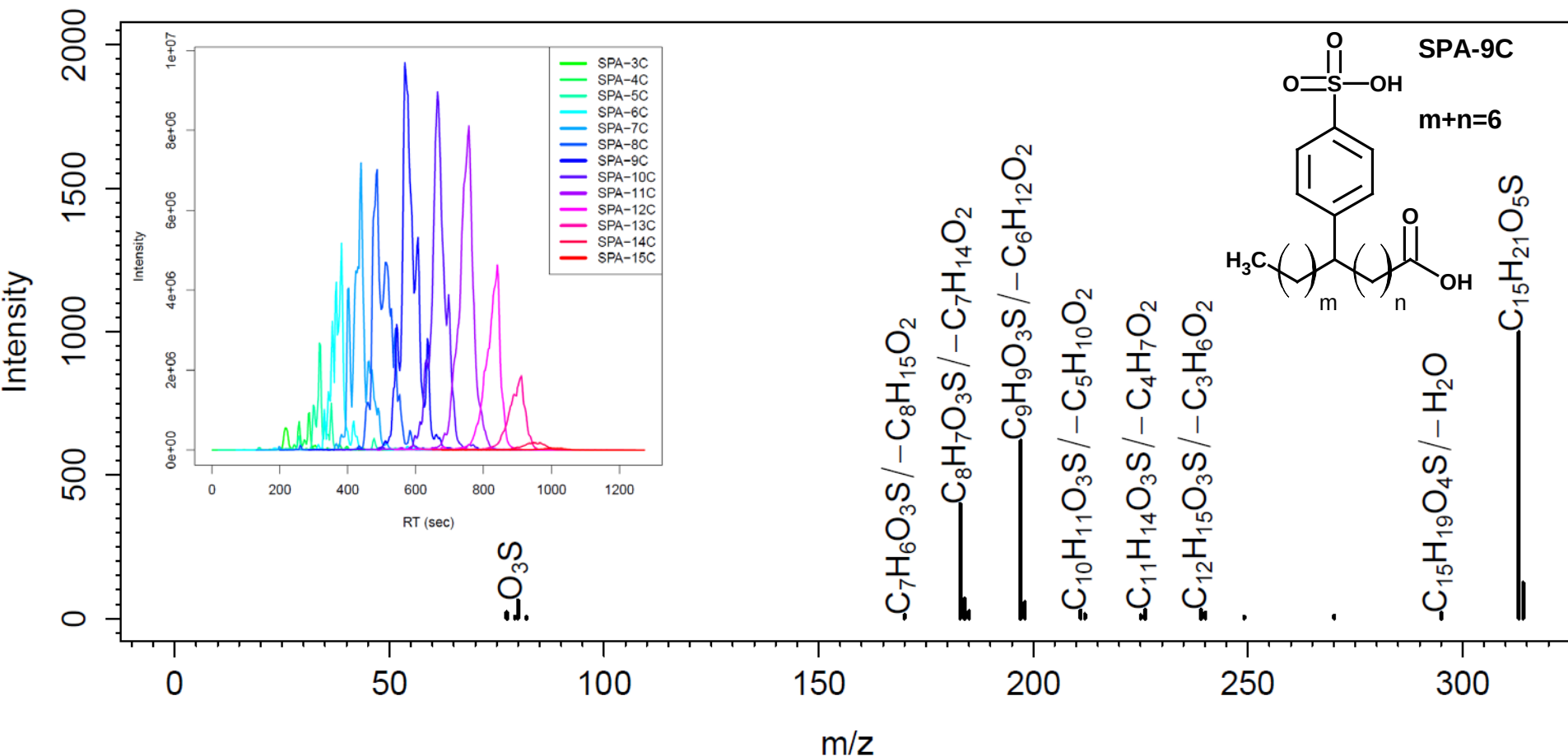


Supporting Evidence for Homologues

Chromatography and MS/MS Annotation

<https://github.com/MassBank/RMassBank/>

RMassBank

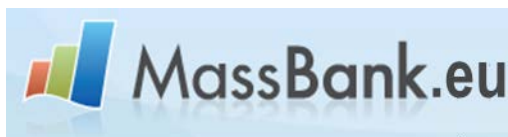


Formulas: <http://sourceforge.net/projects/genform/>

Meringer *et al.* 2011, *MATCH* 65, 259-290

Data: Schymanski *et al.* 2014, *ES&T*, 48:

1811-1818. DOI: 10.1021/es4044374

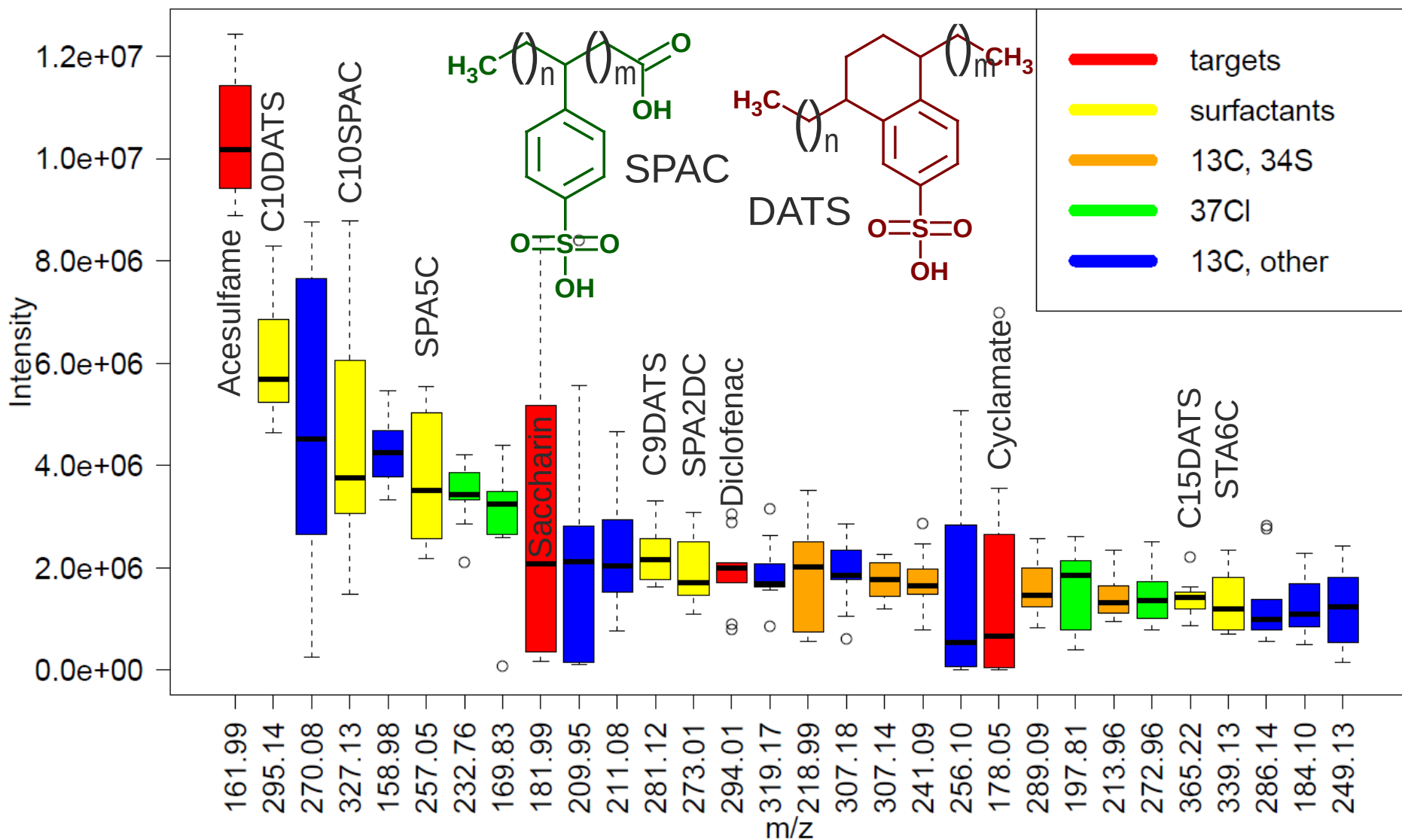


Literature: LIT00034,35

Sample: ETS00002

Standard: ETS00016,17,19,20

Swiss Wastewater: Top 30 Peaks (ESI-)



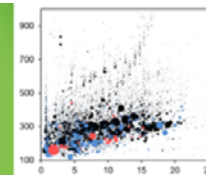
NORMAN Suspect Exchange



o <http://www.norman-network.com/?q=node/236>

NORMAN

Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances



Home | NORMAN Network | Working Groups | Membership | NORMAN Bulletin | Success Stories | Publications | Job opportunities | Contact | Gallery | NORMAN GA meetings

Menu

- Emerging Substances
- DATABASES
- Topics and Activities
- Workshops and Events
- QA/QC Issues
- Glossary

User login

Username *

Password *

[Request new password](#)

[Log in](#)

NORMAN Suspect List Exchange

As part of a series of workshops in September 2014, NORMAN members expressed the need to exchange various lists of substances to improve their suspect screening efforts. An initiative of the 2015 Joint Programme of Activities involved establishing this website as a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring question. All suspect lists currently available are compiled in the table below and are being progressively integrated into the US EPA CompTox Chemistry Dashboard ([website](#), [downloads](#)). The "Link to full list" column below contains an excel or comma-separated file (csv) with all available information, e.g. as provided as supporting information for the publication, while the third column provides a list of the structures as InChIKeys only, which allows suspect searching using MetFrag or other workflows. The fourth column contains references for the data: please cite these references if you use the respective datasets.

Coordination: Emma Schymanski, Eawag; Curation/RTI/toxicity: Reza Aalizadeh & Nikos Thomaidis, Uni. Athens; CompTox: Antony Williams, US EPA; Webmaster: Natalia Glowacka, Environmental Institute; IT: Lubos Cirka, Environmental Institute; Contributors: see below.

If you have any feedback or a list that you would like included, please contact suspects@normandata.eu.

Interactive merged list of all suspect substances (update in progress)

Name and Description	Link to full list	Link to InChIKey list	
Merged NORMAN Suspect List "SusDat"	NORMAN_SusDat_MergedSuspects24052017.xlsx	NORMAN_SusDat_MSready_InChIKeys_24052017.txt	Schymanski et al. 2014. DOI: 10.1021/es4044374
NORMAN Compounds in MassBank	MassBankEU	MassBankEU_compounds_InChIKeys_11042017.txt	See here for an interactive version. Compiled by Reza Aalizadeh, University of Athens, now including RTI and toxicity values.
Eawag Surfactant Suspect List (formulas only)	Surfactant_Suspects_Schymanski_etal_2014.xlsx Surfactant_Suspects_Schymanski_etal_2014.csv	STOFF-IDENT_28102016_InChIKeys.txt	www.massbank.eu Stravs et al. 2012. DOI: 10.1002/jms.3131
Emerging substances	STOFF-IDENT_Content_28102016.xlsx STOFF-IDENT_Content_28102016.csv		The database enables the search for exact masses from target or unknown lists and the automatic use of a Retention Time Index. See: http://bb-x.stoffident.hswt.de - free access after registration
NORMAN Collaborative Trial Targets and Suspects	Targ_Sus_NT-wID_LC_final_31102016.xlsx Targ_Sus_NT-wID_LC_final_31102016.csv Targ_Sus_NT-wID_GC_final_31102016.xlsx Targ_Sus_NT-wID_GC_final_31102016.csv	Targ_Sus_NT-wID_GC_final_InChIKeys_31102016.txt Targ_Sus_NT-wID_LC_final_InChIKeys_31102016.txt	Schymanski et al. 2015. DOI: 10.1007/s00216-015-8681-7

Schymanski, Aalizadeh et al. in prep.

Eawag Surfactant List

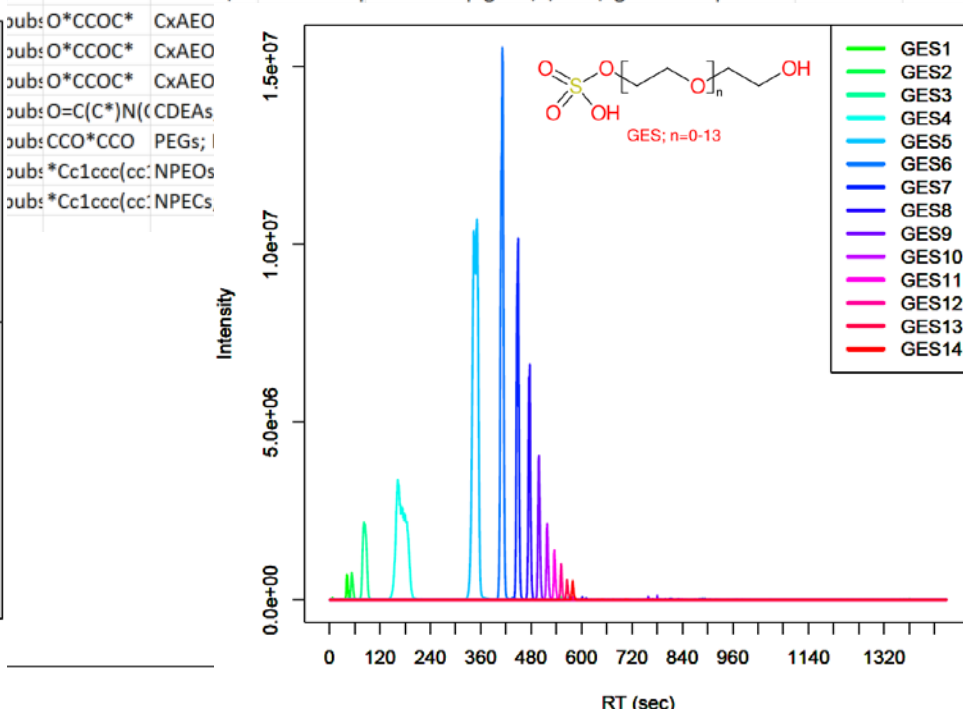
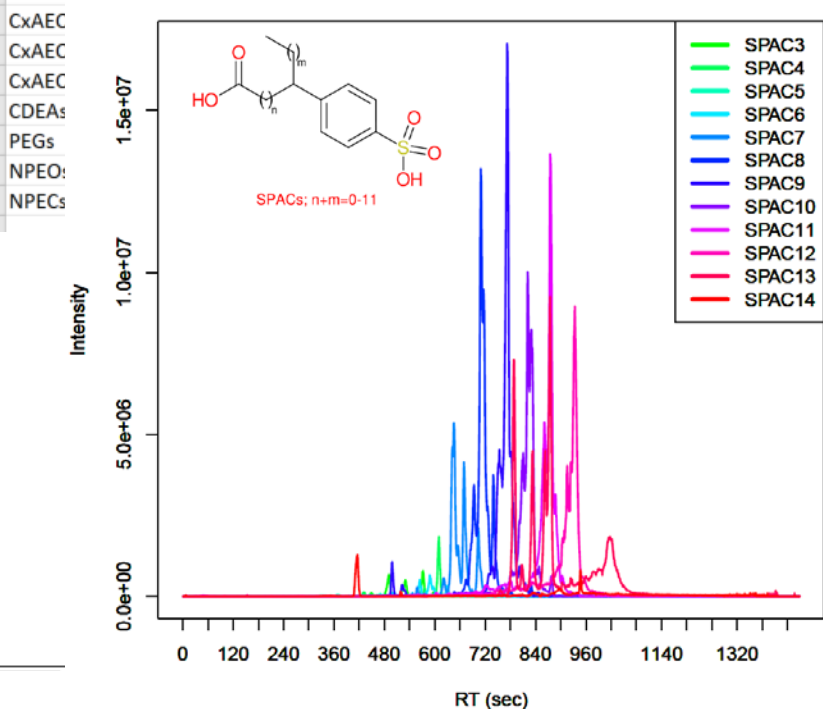
Eawag Surfactant Suspect List (formulas only)	Surfactant_Suspects_Schymanski_etal_2014.xlsx Surfactant_Suspects_Schymanski_etal_2014.csv	Schymanski <i>et al.</i> 2014. DOI: 10.1021/es4044374
---	---	---

SuspectID	Name	Name_ref	Formula	Monoisoto	Adduct_Sta	M+H+	M-H-	Reference	Reference_DOI	Source_ref	Source_DOI
C10-LAS	C10-LAS	C10-LAS_G	C16H26S1O	298.1603	None	299.1675	297.153	Schymansk	dx.doi.org/10.1021/es	Gonzalez_e	dx.doi.org/10.1002/rcm.3527
C11-LAS	C11-LAS	C11-LAS_G	C17H28S1O	312.1759	None	313.1832	311.1686	Schymansk	dx.doi.org/10.1021/es	Gonzalez_e	dx.doi.org/10.1002/rcm.3527
C12-LAS	C12-LAS	C12-LAS_G	C18H30S1O	326.1916	None	327.1988	325.1843	Schymansk	dx.doi.org/10.1021/es	Gonzalez_e	dx.doi.org/10.1002/rcm.3527
C13-LAS	C13-LAS	C13-LAS_G	C19H32S1O	340.2072	None	341.2145	339.1999	Schymansk	dx.doi.org/10.1021/es	Gonzalez_e	dx.doi.org/10.1002/rcm.3527
C14-LAS	C14-LAS	C14-LAS_G	C20H34S1O	354.2229	None	355.2301	353.2156	Schymansk	dx.doi.org/10.1021/es	Gonzalez_e	dx.doi.org/10.1002/rcm.3527
C3-SPC	C3-SPC	C3-SPC_Co	C9H10O5S	230.0249	None	231.0322	229.0176	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C4-SPC	C4-SPC	C4-SPC_Co	C10H12O5S	244.0405	None	245.0478	243.0333	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C5-SPC	C5-SPC	C5-SPC_Co	C11H14O5S	258.0562	None	259.0635	257.0489	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C6-SPC	C6-SPC	C6-SPC_Co	C12H16O5S	272.0718	None	273.0791	271.0646	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C7-SPC	C7-SPC	C7-SPC_Co	C13H18O5S	286.0875	None	287.0948	285.0802	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C8-SPC	C8-SPC	C8-SPC_Co	C14H20O5S	300.1031	None	301.1104	299.0959	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C9-SPC	C9-SPC	C9-SPC_Co	C15H22O5S	314.1188	None	315.1261	313.1115	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C10-SPC	C10-SPC	C10-SPC_C	C16H24O5S	328.1344	None	329.1417	327.1272	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C11-SPC	C11-SPC	C11-SPC_C	C17H26O5S	342.1501	None	343.1574	341.1428	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C12-SPC	C12-SPC	C12-SPC_C	C18H28O5S	356.1657	None	357.173	355.1585	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C13-SPC	C13-SPC	C13-SPC_C	C19H30O5S	370.1814	None	371.1887	369.1741	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C14-SPC	C14-SPC	C14-SPC_C	C20H32O5S	384.197	None	385.2043	383.1898	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
C15-SPC	C15-SPC	C15-SPC_C	C21H34O5S	398.2127	None	399.22	397.2054	Schymansk	dx.doi.org/10.1021/es	Corada-Fer	dx.doi.org/10.1039/c1em10150a
SPA-1DC	SPA-1DC	SPA-1DC_DC	C9H8O7S1	259.9991	None	261.0063	258.9918	Schymansk	dx.doi.org/10.1021/es	DiCorcia_et	dx.doi.org/10.1021/es990596u
SPA-2DC	SPA-2DC	SPA-2DC_DC	C10H10O7S	274.0147	None	275.022	273.0074	Schymansk	dx.doi.org/10.1021/es	DiCorcia_et	dx.doi.org/10.1021/es990596u
SPA-3DC	SPA-3DC	SPA-3DC_DC	C11H12O7S	288.0304	None	289.0376	287.0231	Schymansk	dx.doi.org/10.1021/es	DiCorcia_et	dx.doi.org/10.1021/es990596u
SPA-4DC	SPA-4DC	SPA-4DC_DC	C12H14O7S	302.046	None	303.0533	301.0387	Schymansk	dx.doi.org/10.1021/es	DiCorcia_et	dx.doi.org/10.1021/es990596u

Using Generic Structures for Screening/Linking

o <https://github.com/schymane/RChemMass/>

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Series_ID	GenSMILES_R1toN	nR1toN	ExtraAtom:RDB_R1toN	SeriesName	Source	BabelSmile	Label1	CxSMILES_	CxSMILES_noTitle								
LAS	OS(=O)(=O){C,C}	{9-13}		LAS	http://pubs:OS(=O)(=O)LAS; R1=C,fOS(=O)(=O)OS(=O)(=O)c1ccc(cc1)C(CC)CC Sg:n:13:m:ht,Sg:n:11:n:ht												
SPACs	OS(=O)(=O){C,C}	{1-13}		SPACs	http://pubs:OS(=O)(=O)SPACs; R1= OS(=O)(=O)OS(=O)(=O)c1ccc(cc1)C(CC(=O)O)CC Sg:n:15:m:ht,Sg:n:11:n:ht												
SPADCs	OS(=O)(=O){C,C}	{0-13}		SPADCs	http://pubs:OS(=O)(=O)SPADCs; R1OS(=O)(=O)OS(=O)(=O)c1ccc(cc1)C(CC(=O)O)CC(=O)O Sg:n:15:m:ht,Sg:n:11:n:ht												
DATS	OS(=O)(=O){C,C}	{0-15}		DATS	http://pubs:OS(=O)(=O)DATS; R1= OS(=O)(=O)OS(=O)(=O)c1ccc2c(c1)C(CCC2CC)CC Sg:n:14:m:ht,Sg:n:16:n:ht												
STACs	OS(=O)(=O){C,C}	{1-13}		STACs	http://pubs:OS(=O)(=O)STACs; R1= OS(=O)(=O)OS(=O)(=O)c1ccc2c(c1)C(CCC2CC(=O)O)CC Sg:n:14:m:ht,Sg:n:18:n:ht												
STADCs	OS(=O)(=O){C,C}	{0-13}		STADCs	http://pubs:OS(=O)(=O)STADCs; R1OS(=O)(=O)OS(=O)(=O)c1ccc2c(c1)C(CCC2CC(=O)O)CC(=O)O Sg:n:14:m:ht,Sg:n:18:n:ht												
AS	O=S(=O)(O){C}	{11-15}		AS	http://pubs:O=S(=O)(O)AS; R1=C(1 O=S(=O)(O)O=S(=O)(O)OCC Sg:n:5:n:ht												
C12AES	O=S(O)(=O){CCO,C}	{2-11,11-11}		C12AES	http://pubs:O=S(O)(=O)C12AES; R1O=S(O)(=O)O=S(O)(=O)OCCOCCC Sg:n:5,6,7::ht,Sg:n:9:m:ht												
C13AES	O=S(O)(=O){CCO,C}	{2-11,12-12}		C13AES	http://pubs:O=S(O)(=O)C13AES; R1O=S(O)(=O)O=S(O)(=O)OCCOCCC Sg:n:5,6,7::ht,Sg:n:9:m:ht												
SAS	O=S(O)(=O){C,C}	{9-20}		SAS	http://pubs:O=S(O)(=O)SAS; R1=C,fO=S(O)(=O)O=S(O)(=O)C(CC)CC Sg:n:5:m:ht,Sg:n:7:n:ht												
NPEOSO4s	O=S(=O)(O){OCC,C}	{0-2,9-9}		NPEOSO4s	http://pubs:O=S(=O)(O)NPEOSO4s; O=S(=O)(O)O=S(=O)(O)OCCOc1ccc(cc1)CC Sg:n:5,6,7::ht,Sg:n:14:m:ht												
GES	O=S(=O)(O){OOC}	{1-15}		GES	http://pubs:O=S(=O)(O)GES; R1=O(O=S(=O)(O)O=S(=O)(O)OCCOCCO Sg:n:5,6,7::ht												
CxAEO2	O[R1]CCOC{CCO,C}	{2-2,9-12}		CxAEO2	http://pubs:O*CCOC* CxAEO2; R:OCCOCCC OCCOCCC Sg:n:1,2,3::ht,Sg:n:5:m:ht												



https://comptox.epa.gov/dashboard/chemical_lists/eawagsurf

Schymanski et al/ 2014, ES&T, 48: 1811-1818. DOI: 10.1021/es4044374

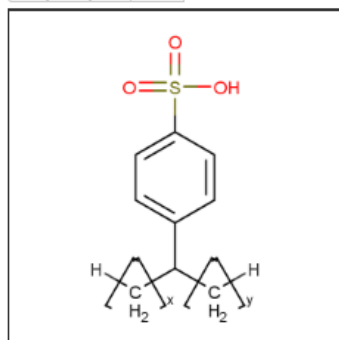
Eawag Surfactant List in CompTox Dashboard

Chemistry Dashboard

Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

Ⓢ Searched by Synonym: Found 1 result for 'Linear alkylbenzene sulfonate'.



Intrinsic Properties

Molecular Formula: $(CH_2)_y(CH_2)_x C_7 H_8 O_3 S$

Average Mass: Not Found

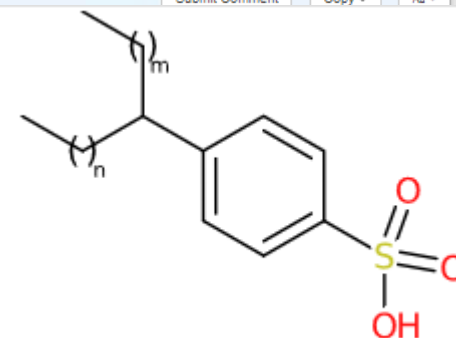
Monoisotopic Mass: Not Found

Structural Identifiers

Linked Substances

Presence in Lists

Record Information



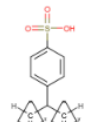
LAS; $n+m=7-10$

[CDK Depict](#)

Related Substances

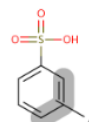
Download

Searched Chemical



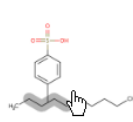
Alkylbenzenesulfonate, linear
42615-29-2

SUCCESSOR: Representative Component



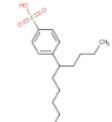
(C10-C18) Alkylbenzenesulfonic acid
68684-22-5

SUCCESSOR: Representative Component



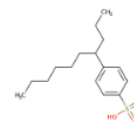
C10-linear alkylbenzenesulfonate
NOCAS_861689

SUCCESSOR: Representative Component



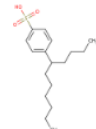
4-(decan-5-yl)benzene-1-sulfonic acid
NOCAS_861146

SUCCESSOR: Representative Component



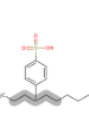
4-(decan-4-yl)benzenesulfonic acid
NOCAS_861333

SUCCESSOR: Representative Component



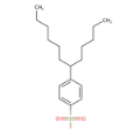
4-(undecan-5-yl)benzene-1-sulfonic acid
NOCAS_861097

SUCCESSOR: Representative Component



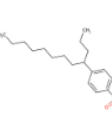
C12-linear alkyl benzene sulfonate
NOCAS_891641

SUCCESSOR: Representative Component



4-(Dodecan-8-yl)benzene-1-sulfonic acid
23003-92-1

SUCCESSOR: Representative Component



4-(dodecan-4-yl)benzene-1-sulfonic acid
NOCAS_862670

SUCCESSOR: General Form

No Chemical
Structure
Associated
with this
Substance

Benzenesulfonic acid, C10-13-alkyl derivs., sodium ...
68411-30-3



https://comptox.epa.gov/dashboard/chemical_lists/eawagsurf



Cross-Linking with Lists in CompTox Dashboard

Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

🔍 Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID3020041'.

Presence in Lists

EPA Hydrofracturing Fluids

Surfactant List Screened in Swiss Wastewater (2014)

Presence in Lists

Surfactant List Screened in Swiss Wastewater (2014)

Surfactant List Screened in Swiss Wastewater (2014)

Record

Quality

C3-C15 Sulfophenyl carboxylates

NOCAS_891722 | DTXSID90891722

🔍 Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID90891722'.

Presence in Lists

MassBank.EU Collection: Special Cases

Surfactant List Screened in Swiss Wastewater (2014)

Record Information

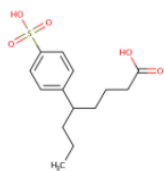
Download as:

Related Chemicals

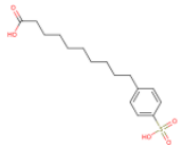
Found 9 chemicals

Download as: TSV Excel SDF

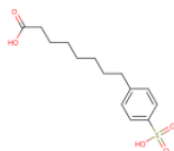
4-(D



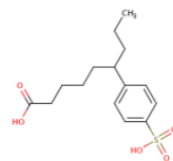
SPA-8C
NOCAS_881094



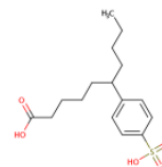
10-(4-sulfophenyl)decanoic acid
NOCAS_891332



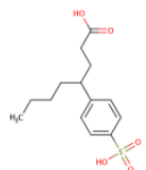
8-(4-sulfophenyl)octanoic acid
NOCAS_891334



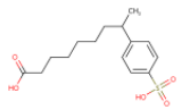
6-(4-sulfophenyl)nonanoic acid
NOCAS_891335



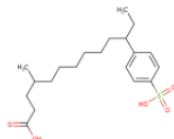
6-(4-sulfophenyl)decanoic acid
NOCAS_891340



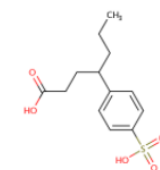
4-(4-sulfophenyl)octanoic acid
NOCAS_891637



8-(4-sulfophenyl)nonanoic acid
NOCAS_891660



4-methyl-11-(4-sulfophenyl)tridecanoic acid
NOCAS_891661



4-(4-sulfophenyl)heptanoic acid
NOCAS_891662

Acknowledgements



RMassBank
nontarget

enviPat Web
enviPick
enviMass



emma.schymanski@uni.lu

Further Information:

www.massbank.eu

<http://www.norman-network.com/?q=node/236>

<https://github.com/MassBank/RMassBank/>

<https://github.com/schymane/RChemMass/>

<https://comptox.epa.gov/dashboard/>

<https://www.wen.uni.lu/lcsb/>



ETH zürich
Department of Biology
Institute of Molecular
Systems Biology



solutions



EU Grant
603437



European (World-)Wide Exchange of Suspects

Tentatively Identified Spectra:

<http://goo.gl/0t7jGp>

Hits in GNPS Massive datasets:

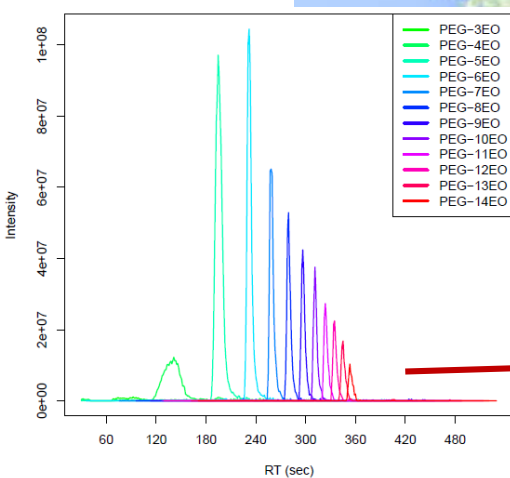
TPs in skin: <http://goo.gl/NmO4tx>

Surfactants: <http://goo.gl/7sY9Pf>

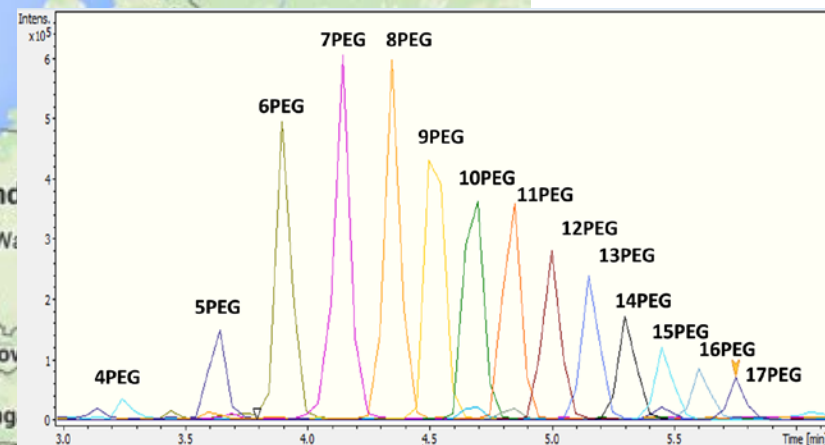
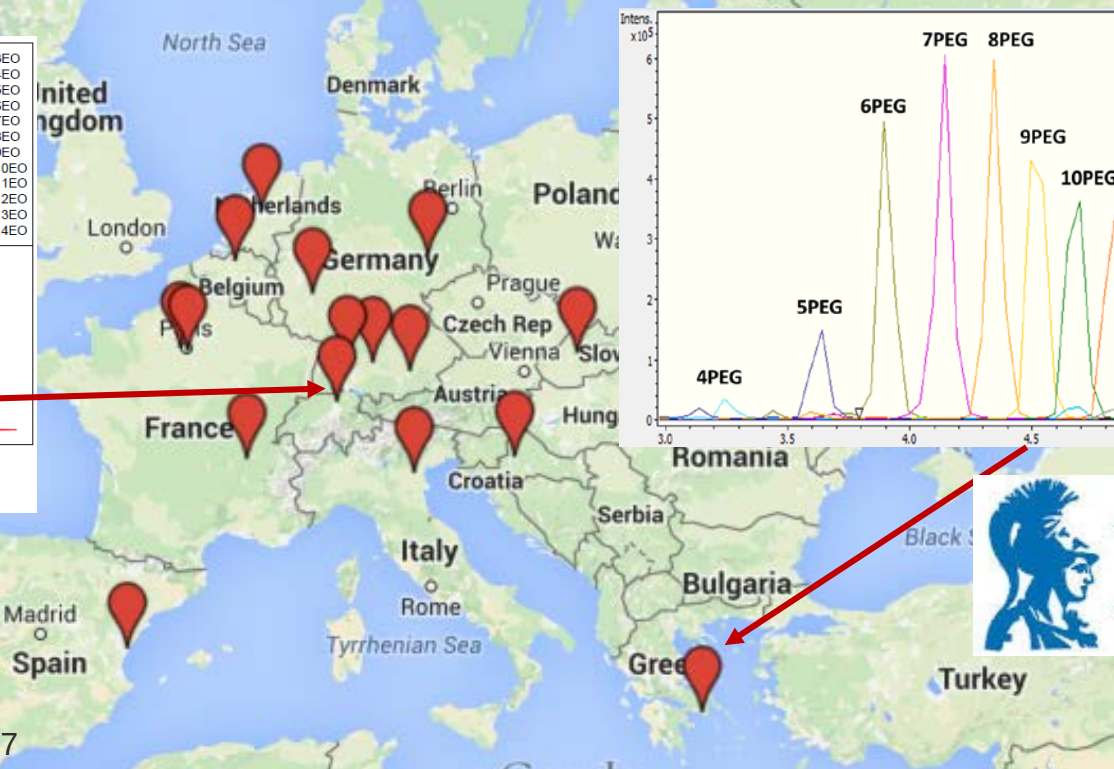
NORMAN Suspect List Exchange:

<http://www.norman-network.com/?q=node/236>

norman news



eawag
aquatic research

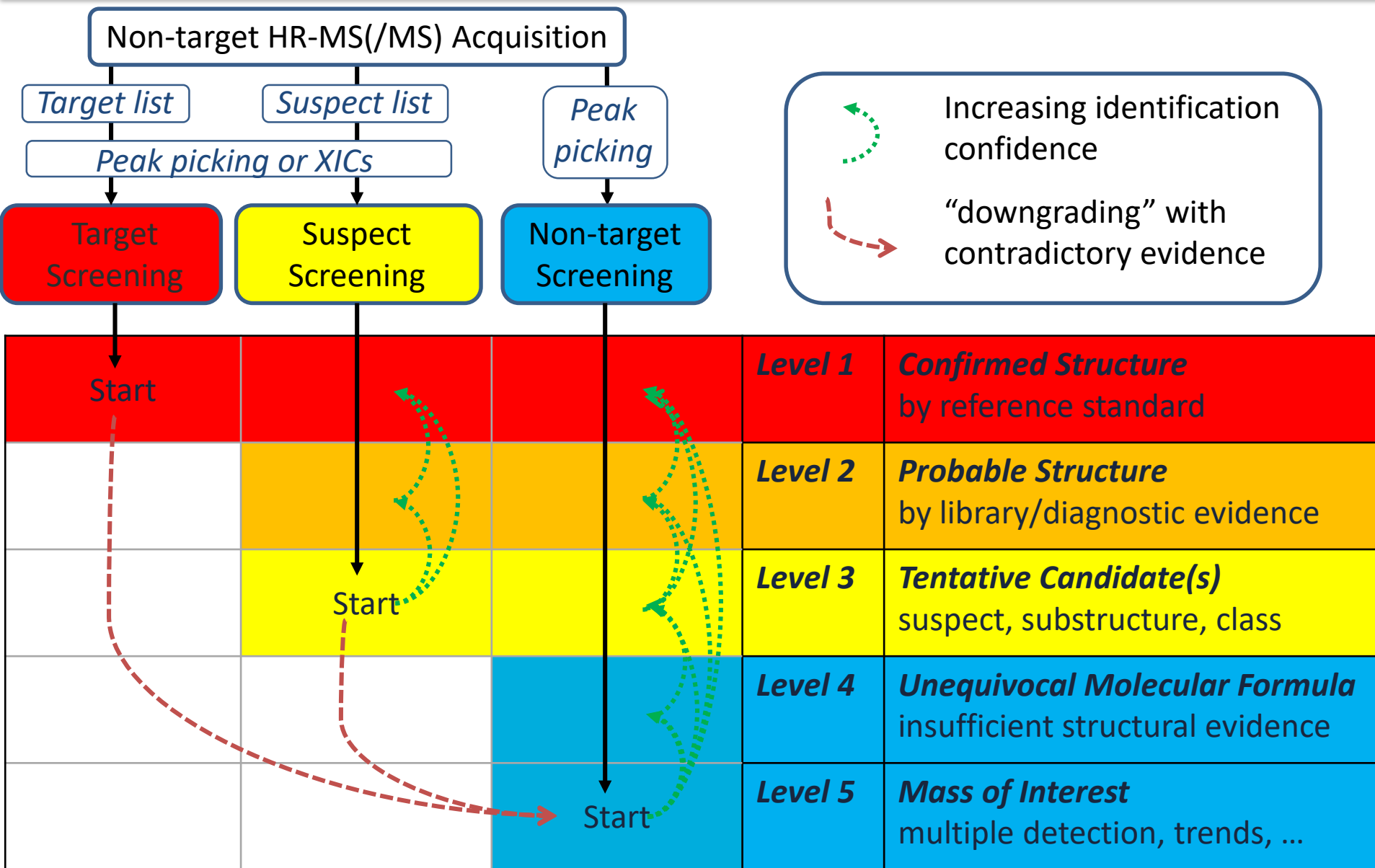


National and Kapodistrian
UNIVERSITY OF ATHENS

Schymanski et al. 2015, ABC,
DOI: 10.1007/s00216-015-8681-7

Schymanski et al. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

2015: European Non-target Screening Trial

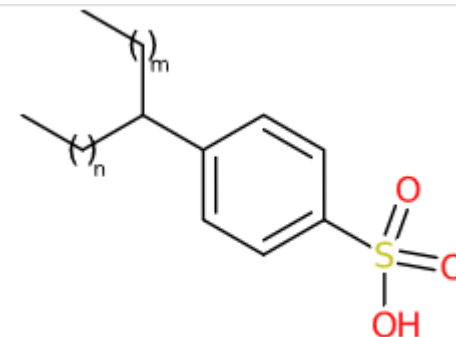


Eawag Surfactant List in CompTox Dashboard

Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

i Searched by Synonym: Found 1 result for 'Linear alkylbenzene sulfonate'.



LAS; $n+m=7-10$

Presence in Lists

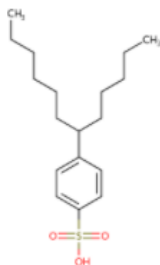
Surfactant List Screened in Swiss Wastewater (2014)

Surfactant List Screened in Swiss Wastewater (2014)

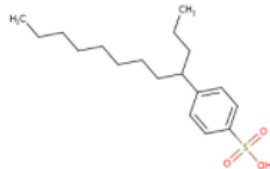
EWAGSURF is a list of surfactants screened in Swiss wastewater effluents as part of a 2014 study. Structures/mixtures are being progressively curated and linked (Schymanski/Williams). Further details in Schymanski et al 2014, DOI: 10.1021/es4044374

[CDK Depict](#)

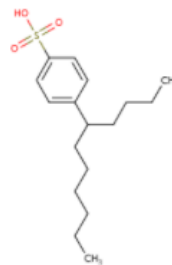
cals



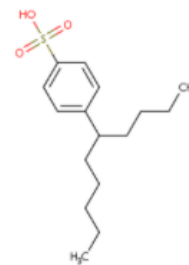
4-(Dodecan-6-yl)benzene-1-sulfon...
23003-92-1



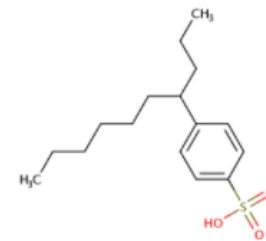
4-(dodecan-4-yl)benzene-1-sulfoni...
NOCAS_862870



C11-LAS
NOCAS_881097



4-(decan-5-yl)benzene-1-sulfonic ...
NOCAS_881146



4-(decan-4-yl)benzenesulfonic acid
NOCAS_891333

<https://www.slideshare.net/AntonyWilliams/>

markush-enumeration-to-manage-mesh-and-manipulate-substances-of-unknown-or-variable-composition

Target, Suspect and Non-Target Screening

